MATHEMATICS

TESTING STATISTICAL HYPOTHESES CONCERNING THE EXPECTATIONS OF TWO INDEPENDENT NORMALS, BOTH WITH VARIANCE ONE. I

BY

WILLEM SCHAAFSMA

(Communicated by Prof. H. FREUDENTHAL at the meeting of October 31, 1970)

0. *Summary*

The use of some objectivistic optimum properties like U.M.P.(D), M.S.(D) and M.S.S.M.P.(D) is illustrated and another property "strongest with respect to Lebesgue-measure" is introduced. The attention is restricted to very simple testing problems for two independent normals. Modifications or generalizations are mentioned sometimes, in order to stress that each of the problems to be considered is the ultimate simplification of certain situations from actual practice. The latter situations are more complicated from a technical point of view (multi-dimensionality, nuisance parameters, non-normality) but the crux is still to decide what is the most appropriate objectivistic optimum property.

Special interest may be requested for (i) the optimum property "strongest w.r. to Lebesgue-measure" (Section 2), (ii) critical remarks concerning maximin and optimal-β in the sections 6 and 7, (iii) a very simple example of an inadmissible likelihood-ratio test (Section 10), (iv) the funny problem of Section 10 where a U.M.P.(D) test exists, D being the class of all size-α tests that are both similar size-α and invariant.

1. *Introduction.* The author is fascinated by the apparently simple problems which arise when the random experiment consists in observing the outcome $x=(x_1, x_2)$ of $X=(X_1, X_2)$ where X_1 and X_2 are independent random variables, X_i having the normal $N(\theta_i, 1)$ distribution. The simplicity of these problems results from the possibility of plotting both the sample point $x=(x_1, x_2)$ and the parameter points $\theta=(\theta_1, \theta_2)$ with respect to the same Cartesian coordinate system in R^2. Of course the sample space is R^2 whereas the parameter space Ω of all possible values of θ will be some subset of R^2. We shall restrict the attention to testing problems in the Neyman-Pearson formulation. The (null-) hypothesis H: $\theta \in \Omega_0$ has to be tested against the alternative (hypothesis) A: $\theta \in \Omega_1$ where $\Omega_0 \cap \Omega_1 = \emptyset$; $\Omega_0 \cup \Omega_1 = \Omega$. For that purpose the statistician will have to construct a test-function $\varphi: R^2 \to [0, 1]$ (φ Borel- or Lebesgue-measurable) such that $\varphi \in \Phi_\alpha$ where Φ_α is the class of all size-α tests (α is a predetermined constant, we always have $\alpha=.05$ in mind). Of course $\varphi \in \Phi_\alpha$ if and only if $E_\theta(\varphi) \leqq \alpha$ for all $\theta \in \Omega_0$.

We shall consider a number of interesting testing problems (they are interesting because they have applications or interpretations in real-life situations). These problems will be formulated by means of the following regions. ω_0 is defined by $\theta_1=\theta_2=0$; ω_1 by $\theta_2=0$; ω_2 by $\theta_2\leqq 0$; ω_3 by $\theta_2>0$; ω_4 by $\theta_1=0$; $\omega_5(\psi)$ by $\theta_1=\theta_2 \operatorname{tg}(\psi)$ and $(\theta_1,\theta_2)\neq(0,0)$; $\omega_6(\psi)$ by $\theta_2>0$ and $|\theta_1|\leqq\theta_2 \operatorname{tg}(\psi)$; $\omega_7(\psi)$ by $\theta_2<0$ and $|\theta_1|\leqq-\theta_2 \operatorname{tg}(\psi)$. Here ψ denotes some fixed angle $0\leqq\psi\leqq\frac{1}{2}\pi$.

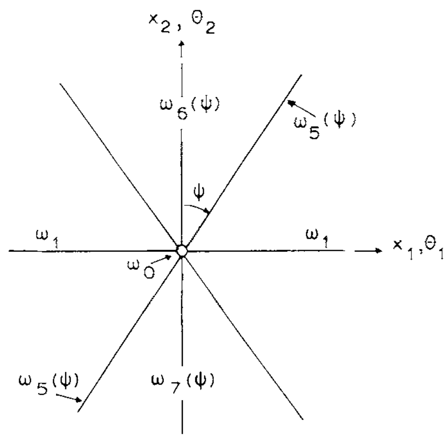

Fig. 1.

The reader is requested to make a drawing of these regions. These drawings are necessary for a good understanding of the rest of this paper. Most of the regions are indicated in fig. 1. but it is much better to have eight separate drawings.

2. *Some objectivistic optimum properties in the Neyman-Pearson theory.* Usually there does not exist a unique U.M.P. (Φ_α) test (of course tests are identified when they are identical a.e. with respect to Lebesgue measure). The objectivist would like to construct complete classes of tests but these results are of limited interest from the practical point of view: the client requires a particular test. A subjectivist might start to argue with his client about "a priori degrees of belief". Objectivists have the opinion that such discussions lead to nothing but indoctrination of the client by the statistician, unless the situation is one with repeated experimentation (see empirical Bayes procedures and other studies which constitute a bridge between the objectivistic and the subjectivistic point of view). Assuming that the situation is one where the client can only be persuaded to select a value of α, and that a U.M.P. (Φ_α) test does not exist, the statistician finds himself in the unfavorable position that he has to propose a particular test though he knows that each proposal will have disadvantages. His only guidelines are the elegance of certain principles and criterions, and the inspection of the power-properties. It is very elegant to restrict the attention to some subclass $D\subset\Phi_\alpha$ by requiring unbiasedness, invariance, similarity (or a combination of these), hoping

that there will exist a U.M.P.(D) test. An interesting example will be given in Section 10. Of course objectivists like to remark that the above-mentioned restrictions may lead to an unreasonable result. In [15] Section 8 for example, there exists a U.M.P. unbiased size-α test which in some situations is less attractive than another very simple S.M.P. size-α test (the sign-test) from an over-all power point of view (of course each U.M.P. unbiased size-α test is admissible).

Next assume that a U.M.P.(D) test does not exist for reasonable subclasses $D \subset \Phi_\alpha$. One might use the generalized likelihood-ratio principle, but the objectivist will prefer to formulate an optimum property for the over-all power function like "maximin size-α (with respect to some partition of Ω_1)" or "optimal-β (with respect to some partition of Ω_1)", and the following optimum properties that are not less attractive for most situations. Let D denote the subclass of Φ_α to which the attention is restricted. Let $\beta_\varphi: \Omega_1 \to [0, 1]$ with $\beta_\varphi(\theta) = E_\theta(\varphi)$ denote the power-function of test φ. Let $\beta_D{}^*: \Omega_1 \to [0, 1]$ with $\beta_D{}^*(\theta) = \sup_{\varphi \in D} \beta_\varphi(\theta)$ denote the envelope power-function with respect to class D. Moreover $\gamma_{\varphi, D} = \beta_D{}^* - \beta_\varphi$ is called the shortcoming of test φ with respect to class D.

Definition 2.1. φ^* is said to be M.S.(D) (most stringent with respect to class D) if and only if (i) $\varphi^* \in D$ and (ii)

$$\sup_{\theta \in \Omega_1} \gamma_{\varphi^*, D}(\theta) = \inf_{\varphi \in D} \sup_{\theta \in \Omega_1} \gamma_{\varphi, D}(\theta). \tag{2.1}$$

In case $D = \Phi_\alpha$, φ^* is called most stringent size-α. This optimum property was introduced by Wald in about 1941. It is often very difficult to construct the M.S.(D) test, unless this test also satisfies other optimum properties like e.g. U.M.P. invariant size-α. For examples of M.S. size-α tests that are not U.M.P. invariant, see [10], [11], [13] and [18]. For testing problems with restricted alternative (see [12]), the criterion M.S.S.M.P.(D) can often be easily satisfied. A test φ is said to be S.M.P.(D) (somewhere most powerful with respect to class D) if and only if (i) $\varphi \in D$ and (ii) $\gamma_{\varphi, D}(\theta) = 0$ for some $\theta \in \Omega_1$. Of course S.M.P.(D) tests are admissible, when $D = \Phi_\alpha$ (or the class of all unbiased size-α tests) and certain uniqueness assumptions are satisfied. Let C denote the class of all S.M.P.(D) tests.

Definition 2.2. φ_0 is said to be M.S.S.M.P.(D) if and only if (i) $\varphi_0 \in C$ and (ii)

$$\sup_{\theta \in \Omega_1} \gamma_{\varphi_0, D}(\theta) = \inf_{\varphi \in C} \sup_{\theta \in \Omega_1} \gamma_{\varphi, D}(\theta). \tag{2.2}$$

Of course $\beta_D{}^* = \beta_C{}^*$ on account of the weak-compactness theorem. Hence φ_0 is M.S.S.M.P.(D) if and only if φ_0 is M.S.(C).

It is obvious that the criterion M.S.S.M.P.(D) may lead to quite unreasonable (though admissible) results. By restricting the attention to C one may throw out the baby with the bathwater; it often seems to be unreasonable to require that $\gamma_{\varphi,D}(\theta)=0$ for some $\theta \in \Omega_1$, one will expect pretty large values of $\gamma_{\varphi,D}$ elsewhere. This obvious disadvantage is also the strength of the criterion M.S.S.M.P.(D) for problems with a restricted alternative (see [12]): S.M.P.(D) tests can often be performed easily while the M.S.(C) = M.S.S.M.P.(D) test φ_0 is determined by *optimal weights which can be determined explicitly* for many problems from actual practice (see [12] Part II).

In the definitions 2.1 and 2.2 we tried to minimize the sup-norm of the shortcoming. This is not a sacred principle at all. One might equally well minimize some other l_p-norm $(1 \leq p < \infty)$ of the shortcoming $\gamma_{\varphi,D}$: $\Omega_1 \to [0, 1]$ (we assume that the attention is restricted to φ's satisfying $\varphi \in D$), though the objectivist will feel uncomfortable because he will have to introduce a certain weight-function and the reasonableness of this depends on what is regarded as a natural parametrization of the class of densities $\{p_\theta(x); \theta \in \Omega_1\}$ over the sample space. Nevertheless the following optimum property seems to be attractive from an objectivistic point of view. We assume that Ω_1 is some two-dimensional subset of R^2 with $\lambda(\Omega_1) > 0$ where λ is Lebesgue-measure; in our applications we have $\lambda(\Omega_1) = \infty$ (Sections 3, 4, 6, 7 and 11 – in Section 11 one has to modify the following theory because Ω_1 is one-dimensional with $\lambda(\Omega_1) = \infty$).

DEFINITION 2.3. φ_1 is called strongest w.r. to λ (and class D) if and only if (i) $\varphi_1 \in \Phi_\alpha$ and (ii)

$$(2.3) \qquad -\infty < \int_{\Omega_1} \gamma_{\varphi_1,D}(\theta)d\lambda(\theta) = \inf_{\varphi \in \Phi_\alpha} \int_{\Omega_1} \gamma_{\varphi,D}(\theta)d\lambda(\theta) < \infty.$$

Class D was introduced in this definition for the sole purpose of remarking that under regularity conditions it does not matter which class D is considered. In the rest of this section we always assume $D = \Phi_\alpha$ with as a result that γ_φ cannot assume negative values and that we can delete D in the notations $\gamma_{\varphi,D}$ and $\beta_D{}^*$.

If $\lambda(\Omega_1) < \infty$ then $\int \beta^* d\lambda < \infty$ and φ_1 is strongest w.r. to λ if and only if φ_1 maximizes $\int \beta_\varphi d\lambda$ or equivalently (Fubini) if and only if φ_1 is M.P. size-α for testing H against the simple alternative that X has the density $\{\lambda(\Omega_1)\}^{-1}\pi(x)$ where

$$(2.4) \qquad \pi(x) = \int_{\Omega_1} p_\theta(x)d\lambda(\theta).$$

If H is simple and p_0 the density under H, then φ_1 is obtained easily by applying the Neyman-Pearson fundamental lemma: φ_1 rejects if and only if the ratio $\varrho = \pi/p_0$ is sufficiently large.

If $\lambda(\Omega_1) = \infty$, then this approach breaks down because it generally turns out that $\int \beta^* d\lambda = \infty$ while $\int \beta_\varphi d\lambda = \infty$ holds for many φ's. Hence there does not exist a *unique* test maximizing $\int \beta_\varphi d\lambda$. Nevertheless there generally (for counter-examples, see Remark 1) does exist a unique test φ_1 satisfying Definition 2.3 while φ_1 again rejects if and only if $\varrho = \pi/p_0$ is sufficiently large (we restrict the attention to a simple hypothesis H). In order to prove this result, we cannot start from the formula

$$\int \gamma_\varphi d\lambda = \int \beta^* d\lambda - \int \beta_\varphi d\lambda$$

because this degenerates into $\infty - \infty$. The trick is now to introduce an auxiliary test φ' for which $\int \gamma_{\varphi'} d\lambda < \infty$, next to write

$$\int \gamma_\varphi d\lambda = \int \gamma_{\varphi'} d\lambda + \int (\beta_{\varphi'} - \beta_\varphi) d\lambda \tag{2.5}$$

and finally to show that the last integral is minimized when φ is the test φ_1.

We work this out in a more abstract context. Suppose the simple hypothesis H has to be tested that X has the probability distribution P_0 over the outcome space.

The composite alternative A states that X has a distribution P_θ out of the class $\{P_\theta; \theta \in \Omega_1\}$. Ω_1 is a subset of some R^p such that Lebesgue-measure can be defined over Ω_1 and $\lambda(\Omega_1) > 0$. We assume that P_0 and all P_θ's are absolutely continuous with respect to some σ-finite measure μ over $\mathscr{X}$ (measurability conditions are disregarded). The corresponding (non-negative regular version of the) Radon-Nikodym derivatives are denoted by p_0 and p_θ.

THEOREM 1. *Assuming* $0 < \alpha < 1$, $0 < p_0(x) < \infty$ *and* $0 < \pi(x) < \infty$ *for all* $x \in \mathscr{X}$ *(π being defined in* (2.4)*), the following holds.*

(i) *Existence. There exists a test* φ_1 *and a constant* $k \in (0, \infty)$ *such that*

$$E_0(\varphi_1) = \int_{\mathscr{X}} \varphi_1 \, dP_0 = \alpha \tag{2.6}$$

and, using the notation $\varrho = \pi / p_0$,

$$\varphi_1(x) = \begin{cases} 0 \text{ when } \varrho(x) < k \\ 1 \text{ when } \varrho(x) > k \end{cases} \tag{2.7}$$

(ii) *Sufficiency. If there exists a size-*α *test* φ' *with* $\int \gamma_{\varphi'} d\lambda < \infty$, *then each test* φ_1 *satisfying* (2.6) *and* (2.7) *is strongest w.r. to* λ.

(iii) *Necessity. If* φ *is strongest w.r. to* λ *then* $E_0(\varphi) = \alpha$ *and there exists a constant* k *such that* $\varphi(x) = \varphi_1(x)$ *for almost* (μ) *all* x *with* $\varrho(x) \neq k$.

PROOF. The formulation of Theorem 1 closely resembles the formulation of the Neyman-Pearson fundamental lemma in LEHMANN [9]. The proof is also similar, but there appear some peculiar difficulties.

(i) See [9] p. 65; it also follows that, under certain regularity assumptions, the constant k is uniquely determined. Always $\infty > k > 0$.

(ii) On account of (2.5) we try to find $\varphi \in \Phi_\alpha$ such that

$$\int_{\Omega_1} (\beta_\varphi - \beta_{\varphi'}) \, d\lambda = \int_{\mathscr{X}} (\varphi - \varphi') \pi \, d\mu \tag{2.8}$$

(Fubini) is maximized. We must show that φ_1 does the job. For that purpose, suppose $\varphi \in \Phi_\alpha$ and φ_1 is determined by (2.6) and (2.7). Then, writing formally,

$$\text{(2.9)}\quad \left\{ \begin{aligned} &\int_{\mathscr{X}} (\varphi_1-\varphi')\,\pi d\mu - \int_{\mathscr{X}} (\varphi-\varphi')\,\pi d\mu = \int_{\{x;\varphi_1(x)>\varphi(x)\}} (\varphi_1-\varphi)\,\pi d\mu + \\ &+ \int_{\{x;\varphi_1(x)<\varphi(x)\}} (\varphi_1-\varphi)\,\pi d\mu \overset{(2.7)}{\geqq} k \int (\varphi_1-\varphi)\,p_0 d\mu = k\{E_0(\varphi_1)-E_0(\varphi)\} \overset{(2.6)}{\geqq} 0. \end{aligned} \right.$$

This computation makes sense when all integrals appearing are finite. This is insured by the existence of a test φ' with $\int \gamma_{\varphi'} d\lambda < \infty$.

In order to show that $\int (\varphi_1-\varphi')\pi d\mu$ is finite, consider first the negative part

$$\int_{\mathscr{X}} (\varphi_1-\varphi')^{-}\pi d\mu = \int_{\{x;\varphi_1(x)<\varphi'(x)\}} (\varphi'-\varphi_1)\,\pi d\mu \overset{(2.7)}{\leqq} k\, E_0(\varphi').$$

Thus our integral is finite or $+\infty$. In the latter case we find (see (2.5) and (2.8))

$$\int_{\Omega_1} \gamma_{\varphi_1} d\lambda = \int_{\Omega_1} \gamma_{\varphi'} d\lambda - \int (\varphi_1-\varphi')\,\pi d\mu = -\infty$$

which is a contradiction because $\int \gamma_{\varphi_1} d\lambda \geqq 0$ on account of $\varphi_1 \in \Phi_\alpha$.

With respect to the finiteness of $\int (\varphi-\varphi')\pi d\mu$, we remark that we can simply restrict the attention to φ's for which this integral is finite.

This completes the proof of (ii).

(iii) Now the conditions of (ii) are satisfied; φ and φ' both are strongest w.r. to λ. The inequalities in (2.9) all are equalities. But $k>0$. Hence $E_0(\varphi)=\alpha$ and

$$\int_{\{x;\varphi_1(x)>\varphi(x)\}} (\varphi_1-\varphi)\,\pi d\mu = k \int_{\{x;\varphi_1(x)>\varphi(x)\}} (\varphi_1-\varphi)\,p_0 d\mu$$

with as a result that

$$\mu(\{x;\,\varphi_1(x)>\varphi(x)\} \cap \{x;\,\pi(x)\neq kp_0(x)\})=0.$$

Similarly

$$\mu(\{x;\,\varphi_1(x)<\varphi(x)\} \cap \{x;\,\pi(x)\neq kp_0(x)\})=0.$$

This completes the proof.

Remark 1. For testing a simple hypothesis, one can often show that the classical test that is optimal for the unrestricted alternative, can be used as the auxiliary test φ'.

When the hypothesis is $\infty-$ composite, then usually $\int \gamma_\varphi d\lambda = \infty$ for all $\varphi \in \Phi_\alpha$ with as a result that Definition 2.3 cannot be satisfied. This establishes an important disadvantage of the criterion "strongest w.r. to λ" compared with "M.S. size-α": a M.S. size-α test exists under regularity conditions, but Definition 2.3 cannot generally be satisfied for composite H.

This is illustrated by means of the following example. $X=(X_1, X_2, X_3)$; X_i's independent $N(\theta_i, 1)$. Hypothesis H: $\theta_1=\theta_2=0$ (θ_3 unspecified); Alternative A: $(\theta_1, \theta_2) \in \Omega_1'$ (θ_3 unspecified) where Ω_1' is some subset of R^2. Let $\varphi_1: R^2 \to [0, 1]$ be the optimal test for the two-dimensional problem which is obtained when X_3 and θ_3 are deleted. One would like to propose the test φ: $R^3 \to [0, 1]$ with $\varphi(x_1, x_2, x_3) = \varphi_1(x_1, x_2)$ for the three-dimensional problem. But obviously $\int \gamma_\varphi d\lambda = \infty$.

Remark 2. It is interesting to remark that in the case of an unrestricted parameter-space we often have $\pi(x) = c$ for all $x \in \mathscr{X}$. In that case φ_1 rejects for outcomes x with $p_0(x)$ sufficiently small. Example: let X have the binomial $B(n, \theta)$ distribution over $\mathscr{X} = \{0, 1, \ldots, n\}$; μ is counting measure. H: $\theta = \varrho$ and A: $\theta \neq \varrho$ where ϱ is some fixed number, say $\varrho = \frac{1}{4}$. We get $\pi(x) = \int \binom{n}{x} \theta^x (1-\theta)^{n-x} d\theta = 1/(n+1)$. Test φ_1 rejects for x whose probability $p_\theta(x)$ is sufficiently small. This is Fisher's method of constructing a critical region by collecting the outcomes with the smallest probabilities. It now follows that this method leads to a test φ_1 that is strongest w.r. to λ (of course φ_1 is not unbiased).

Remark 3. (the "fundamental dilemma" in the objectivistic approach to statistical inference). For many of the simple problems to be considered, different objectivistic optimum properties are satisfied by different tests. We are fascinated by these problems because they constitute examples of this fundamental dilemma which are not obscured by technical difficulties.

One will have "to compare the optimum properties" with respect to the specific situation under consideration. We distinguish three points of view: (i) the elegance, (ii) the resulting power properties, (iii) the manageability of the test obtained. Unfortunately, a confrontation of these aspects may easily lead to a conflict.

With respect to (i), we believe that many objectivists will agree upon the following guidelines. Global optimum properties, like U.M.P.(D), M.S.(D), strongest w.r. to λ, M.S.S.M.P.(D), maximin, optimal (β), are more elegant than local optimum properties which are more elegant than for example the property of being obtained by means of the (generalized) likelihood-ratio principle. The latter principle is not an optimum property at all; it is a principle to derive a test statistic which principle is formulated without reference to the power properties. Among the global optimum properties one will consider U.M.P.(D) to be most elegant, at least when D is a nice subclass of Φ_α which is obtained by requiring unbiasedness, invariance and (or) similarity. Of course U.M.P.(D) is more elegant than U.M.P.(D′) when $D \supset D'$. Next suppose that U.M.P.(D) cannot be satisfied. Then one will prefer M.S.(D) and strongest w.r. to λ over M.S.S.M.P.(D) from the point of view of elegance. With respect to maximin and optimal (β) we remark that these properties are often satisfied with respect to a certain distance-function or partitioning of Ω_1 and that it is often shown that a particular S.M.P.(D) test is (uniformly) maximin or optimal (β)

with respect to such a partitioning. In such situations the criterion M.S.S.M.P.(D) is obviously more elegant.

With respect to (ii), (the power properties), it is very well possible that from a common sense point of view (see [13]), for example, the power properties of a M.S.S.M.P.(D) test are preferred over those of a M.S.(D) test. Or that a particular S.M.P. size-α test (which is not unbiased) is preferred over a U.M.P. unbiased size-α test ([15] Section 8).

With respect to (iii), (the manageability of the test obtained), it often happens that less elegant optimum properties are satisfied by tests that can be managed better. In [12], for example, easily applicable M.S.S.M.P.(D) tests are constructed for very intricate problems from actual practice, for which it is very difficult to construct the likelihood-ratio size-α tests and practically impossible to satisfy the more elegant optimum properties M.S.(D) and strongest w.r. λ.

We shall now consider the problems that were introduced at the end of Section 1.

3. $\mathbf{\Omega_0=\omega_0;\ \Omega_1=R^2-\omega_0}$. The classical χ^2 test φ' that rejects if and only if $x_1{}^2+x_2{}^2 \geqq \chi^2_{2;\alpha}$ is U.M.P. invariant size-α, M.S. size-α (see [9]) and strongest w.r. λ. For the last property, see Theorem 1; $\pi(x)=1$ for all x; $\varrho(x)=2\pi \exp\{\frac{1}{2}(x_1{}^2+x_2{}^2)\}$, hence $\varphi_1(x)=1$ if and only if $x_1{}^2+x_2{}^2$ is sufficiently large.

4. $\mathbf{\Omega_0=\omega_0;\ \Omega_1=R^2-\omega_1}$. Compare with Section 3. The invariance considerations are no longer of application, but the χ^2 test φ' is still M.S. size-α and strongest w.r. λ. This test will not have any reasonable competitor from an objectivistic point of view of over-all power properties.

Remark. The formulation Ω_1: $\theta_2 \neq 0$ strongly suggests that the test which rejects if and only if $|x_2|>u_{\frac{1}{2}\alpha}$ is very reasonable. Of course this test is admissible; it is even S.M.P. unbiased size-α but it is a very bad test from an over-all point of view, with the maximum shortcoming equal to 1-α (higher dimensional cases in [12] Sections 2.13 and 2.14.)

A problem from practice. Suppose one observes k independent random variables S_i having binomial $B(n_i, p_i)$ distributions ($i=1, \ldots, k$; n_i's known, p_i's unknown successprobabilities). Hemelrijk and Van Eeden considered testing H: $p_1=\ldots=p_k$ of homogeneity of the probabilities against the alternative A: $\sum(2i-k-1)p_i \neq 0$. They proposed a two-sided test based on the test-statistic $\sum(2i-k-1)n_i^{-1}S_i$. From the discussions above it follows that this is a bad test for testing H against A: the classical χ^2-test with $k-1$ degrees of freedom is approximately M.S. size-α because it has this property for the unrestricted alternative (this contradicts Remark 3 in [8] p. 195 where it is surmised that the power properties of the χ^2-test are worse).

5. $\mathbf{\Omega_0=\omega_1}$ or $\mathbf{\Omega_0=\omega_2;\ \Omega_1=\omega_3}$. The test which rejects if and only if $x_2>u_\alpha$ is U.M.P. size-α. This is shown as follows. For $(\theta_1, \theta_2) \in \Omega_1$ introduce

$(\theta_1, 0)$ as the least-favorable parameterpoint in Ω_0. The above-mentioned test is M.P. for testing $(\theta_1, 0)$ against (θ_1, θ_2) (this follows from the Neyman-Pearson fundamental lemma). But the test belongs to the class of size-α tests for testing the composite hypothesis H: $\theta \in \Omega_0$. Hence M.P. size-α for H against the simple alternative $(\theta_1, \theta_2) \in \Omega_1$. Hence U.M.P. size-$\alpha$.

REMARK. In multidimensional situations (see for example FERGUSON's beautiful book [7] Section 5.4 Problem 6), one often starts by applying the general method for constructing U.M.P. unbiased size-α tests for multiparameter exponential families. In one-sided situations one ought to be careful and try to show that the U.M.P. unbiased test is also U.M.P. size-α, for the latter criteron is much more compelling. In [15] Section 8 a problem from practice is described where in certain situations a U.M.P. unbiased test exists which is less attractive from an over-all point of view than the sign test which obviously then will not be unbiased (the sign test is one of the S.M.P. size-α tests).

6. $\boldsymbol{\Omega_0 = \omega_0}$; $\boldsymbol{\Omega_1 = \omega_3 = \omega_6(\frac{1}{2}\pi)}$. This is a one-sided analogue of the problem of Section 4 and a limiting situation for Section 7. The situation is less clear than in Section 4 because the M.S. size-α test φ^* is unknown. In [14] the problem to obtain φ^* was described as a continuous linear programming problem that might be solved iteratively by means of successive discretizations; but the numerical task seems to be forbidding. The idea behind the results in [11], [13] and [18] is that one could guess that the solution of the dual problem (the least favorable distribution over the alternative) was of a certain kind. This guess turned out to be correct in certain situations but wrong in others. For the problem under consideration guesses can be made but we are afraid that, after performing the corresponding enormous computational task, they may turn out to be wrong.

STRONGEST WITH RESPECT TO LEBESGUE-MEASURE. Apply Theorem 1. Let Φ denote the cumulative distribution function of the normal $N(0, 1)$. Then

$$\pi(x) = \Phi(x_2); \ \varrho(x) = (2\pi)^{\frac{1}{2}}\Phi(x_2) \exp\{\tfrac{1}{2}(x_1{}^2 + x_2{}^2)$$

and the test φ_1 will reject for large outcomes of

$$T = X_1{}^2 + X_2{}^2 + 2\ {}^e\!\log\{\Phi(X_2)\}. \tag{6.1}$$

One can show that $x^2 + 2\ {}^e\!\log\{\Phi(x)\}$ is a strictly increasing (even convex: positive second derivative) function of x over the real line, climbing from $-\infty$ to $+\infty$. The critical region of φ_1 will be sketched in fig. 2; in fig. 3 power-properties will be characterized.

LIKELIHOOD-RATIO. The likelihood-ratio test φ_2 can be obtained easily and is sketched in fig. 2 (multi-dimensional situations have been considered in [12], Section 2.15).

Next we consider three other approaches, each providing the test φ_0 which rejects if and only if $x_2 \geqq u_\alpha$. Though this test is admissible (it is the unique (a.e. λ) U.M.P. test against $\omega_6(0)$), it will follow from the power comparison at the end of this section that φ_0 is a bad test. This could also be anticipated from Section 4. Thus the following three approaches may lead to unsatisfactory results.

MOST STRINGENT SOMEWHERE MOST POWERFUL. The test that rejects for large values of $\theta_1 X_1 + \theta_2 X_2$ is M.P. against the simple alternative (θ_1, θ_2). Thus the class of all S.M.P. size-α tests is obtained easily and consists of the tests that reject for large outcomes of the corresponding test-statistic $w_1 X_1 + w_2 X_2 (w_2 > 0)$. It is seen easily that each test of this kind has the maximum shortcoming equal to 1 except for the test φ_0 which is obtained when $w_1 = 0$. This test φ_0 has the maximum shortcoming equal to 1-α and hence this is the unique M.S.S.M.P. size-α test.

UNIFORMLY MAXIMIN. Obviously each size-α test has the minimum power over Ω_1 not larger than α. Each unbiased size-α test has maximin power, and consequently the maximin criterion is not of direct application. One might argue as follows. The formulation $\Omega_1 = \{(\theta_1, \theta_2);\, \theta_2 > 0\}$ suggests that it is natural to consider an indifference zone $\Omega_I(\varrho) = \{(\theta_1, \theta_2);\, 0 < \theta_2 < \varrho\}$ or, equivalently, subalternatives of the form $A_\varrho : \theta \in \omega_\varrho = \{(\theta_1, \theta_2);\, \theta_2 \geqq \varrho\}$. Now it is seen easily that φ_0 is the unique maximin size-α test for testing $H : \theta \in \Omega_0 = \omega_0$ against A_ϱ. But φ_0 does not depend on ϱ and hence φ_0 is "uniformly maximin" with respect to the above-mentioned class of subalternatives. Thus a S.M.P. test is shown to have some maximin property. Similar results are proved by Doksum a.o. (see for example [5]) for intricate nonparametric problems where it seems to be very difficult to find the M.S.S.M.P. test. Nevertheless the example shows that the criterion uniformly maximin is not very satisfactory because it depends heavily on the class of subalternatives considered. If one would have defined A_ϱ': $\theta \in \omega_\varrho' = \{(\theta_1, \theta_2);\, \theta_2 > 0,\, \theta_1^2 + \theta_2^2 \geqq \varrho\}$ then φ_0 is no longer maximin.

OPTIMAL-β FOR EACH $\beta > \alpha$. This approach is similar to the maximin approach. Let $\varrho(\beta)$ be the smallest value of ϱ such that $E_\theta(\varphi_0) \geqq \beta$ for all $\theta \in A_\varrho$. It is seen easily that each other size-α test φ ($\varphi = \varphi_0$ a.e. λ does not hold) satisfies $E_\theta(\varphi) < \beta$ for some $\theta \in A_{\varrho(\beta)}$. This establishes a certain optimum property of φ_0: φ_0 is optimal-β for each $\beta > \alpha$ with respect to the class of subalternatives $\{A_\varrho;\, \varrho > 0\}$ (see DAVIES [4]).

A POWER-COMPARISON. For $\alpha = .05$ the critical regions defined by φ_0, φ_1 and φ_2 are drawn in fig. 2. Let

$$\gamma_\varphi(\psi) = \sup_{\varrho > 0} \gamma_\varphi(\varrho \sin \psi, \varrho \cos \psi)$$

denote the maximum of the shortcoming of test φ over the half-line l in Ω_1 that makes an angle ψ with the positive x_2, θ_2-axis (see fig. 2). A very good idea of the power-properties of the three tests φ_0, φ_1 and φ_2 can be obtained by plotting the corresponding graphs of $\gamma_\varphi(\psi)$ as a function of ψ (fig. 3).

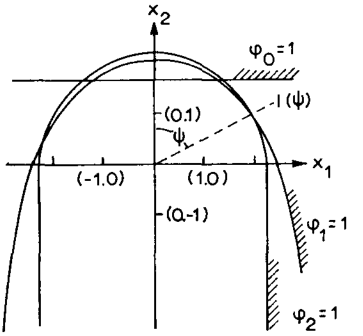

Fig. 2.

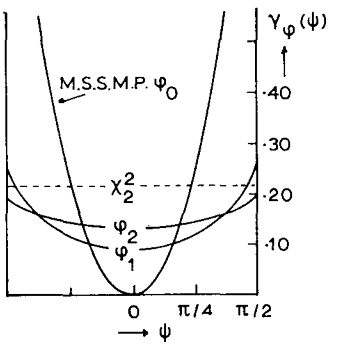

Fig. 3.

The classical χ^2 test which rejects for $X_1^2+X_2^2$ sufficiently large, is also mentioned in fig. 3. It turns out that the M.S.S.M.P. test φ_0 is very bad. By restricting the attention to the S.M.P. tests, we have thrown out the baby with the bathwater. Both φ_1 and φ_2 are attractive. We would prefer φ_1 from an over-all point of view.

A PROBLEM FROM PRACTICE. Consider the one-sided analogue of the problem at the end of Section 4 where now H: $p_1=p_2=\ldots=p_k$ has to be tested against $A: \sum(2i-k-1)p_i>0$. The test rejecting for large values of $\sum(2i-k-1)n_i^{-1}S_i$ plays a part similar to that of φ_0. Thus this will be a bad test though one might be able to show that the test is "approximately M.S.S.M.P. size-α". It is much better to apply the likelihood-ratio principle (see [12] Section 2.15) or to replace the problem by an asymptotically equivalent problem that is formulated in terms of normally distributed random variables, in order to propose to use a test that is good for the latter problem (see [12] Section 2.11).

7. $\mathbf{\Omega_0=\omega_0;\ \Omega_1=\omega_6(\psi)}$. This is the simplest situation of a non-degenerate testing problem where the alternative is restricted by a number of inequalities; here the inequalities

$$\theta_1\cos\psi+\theta_2\sin\psi\geqq 0,\quad -\theta_1\cos\psi+\theta_2\sin\psi\geqq 0$$

(see [12]).

Especially the case $\psi=\pi/4$, where Ω_1 is a quadrant, has been considered carefully. In [12] Ch. 4 the M.S.S.M.P. test φ_0 and the likelihood-ratio test were considered. In [11] and [18] VAN ZWET and OOSTERHOFF made a very careful comparison of the power properties of the M.S. size-α test φ^*, the M.S.S.M.P. size-α test φ_0, the likelihood-ratio size-α test φ_2

and some other tests based on attractive test-statistics among which Fisher's omnibus test φ_3 which rejects for large values of

$$\{1-\Phi(Y_1)\}\,\{1-\Phi(Y_2)\} \tag{7.1}$$

where $Y_1=(X_1+X_2)/\sqrt{2}$ and $Y_2=(X_2-X_1)/\sqrt{2}$.

The surprising result of their power-comparison is that this test of Fisher is extremely attractive from the point of view of over-all power, notwithstanding the "optimality" of φ^* and φ_0. This result raises curiosity with respect to the test φ_1 that is strongest w.r. λ.

In [13] a comparison was made of φ_0 and φ^* for arbitrary values of ψ (satisfying the restriction that the M.S. test φ^* could be obtained). The result was that for ψ "sufficiently small" neither of these two tests provides a worthwhile improvement upon the other. It was boldly extrapolated that then φ_0 cannot be improved upon to a worthwhile extent by any other test. OOSTERHOFF's work [11] shows that this extrapolation is not correct. For $\psi=\pi/4$ the power properties of φ_3 seem to be nicer than those of either φ_0 or φ^*.

MOST STRINGENT SIZE-α. For obtaining φ^* and its power see [13]. There φ^* was obtained for $\psi\leqq\psi_0^{(cr)}(\alpha)$ where $\psi_0^{(cr)}(\alpha)$ is a certain critical angle.

STRONGEST W.R. TO λ. To obtain φ_1 is a straightforward but computationally intricate application of Theorem 1. Only for $\psi=\frac{1}{2}\pi$ (Section 6) and $\psi=\pi/4$ (the van Zwet-Oosterhoff case of a quadrant) simplifications are possible. Thus consider the $\psi=\pi/4$ case. It is convenient to introduce new axes along the edges of Ω_1. Thus we consider the problem with $\Omega_0'=\omega_0$; $\Omega_1'=\{(\theta_1,\theta_2);\ \theta_1>0,\ \theta>0\}$. Apply Theorem 1. $\pi(x)=\Phi(x_1)\Phi(x_2)$. Test φ_1 rejects for large outcomes of

$$\sum_{i=1}^{2}\{X_i^2+2\,{}^{e}\!\log\Phi(X_i)\} \tag{7.2}$$

(see the discussion after formula (6.1)). Of course by applying a simple rotation we obtain the test φ_1 for our original problem.

LIKELIHOOD-RATIO SIZE-α. Test φ_2 is obtained easily. The critical region is bounded by two half-lines and the part of a circle inside $\Omega_1=\omega_6(\psi)$.

MOST STRINGENT SOMEWHERE MOST POWERFUL SIZE-α. Test φ_0 which rejects if and only if $x_2\geqq u_\alpha$ is easily seen to be the unique M.S.S.M.P. size-α test. Obviously the criterion M.S.S.M.P.(D) is attractive from the practical point of view: one does not need new tables. It has been shown in [12] that many testing problems where the alternative is restricted by inequalities, can be attacked by means of criterions M.S.S.M.P.(D) and that the tests obtained can be applied easily. Unfortunately the power properties of the M.S.S.M.P.(D) test φ_0 may be very bad as we have seen in Section 6 for the $\psi=\frac{1}{2}\pi$ case. We have conjectured in [12] and [13] that one cannot improve upon the power-properties of φ_0 to a worthwhile extent if ψ is "sufficiently small". Then, of course, if ψ is "sufficiently small", φ_0 is the most attractive test because of its simplicity. The basic problem is to find out which values of ψ are "sufficiently small". OOSTERHOFF's work [11] shows that we were too optimistic in [12] and [13]: for $\psi = \pi/4$ and $\alpha = .05$ there exist tests like Fisher's test φ_3 which have nicer power properties than φ_0 from a common-sense over-all point of view.

UNIFORMLY MAXIMIN AND OPTIMAL-β. These criterions are satisfied by φ_0, if we consider the sub-alternatives A_ϱ: $\theta \in \omega_\varrho = \{\theta; \theta = (\theta_1, \theta_2) \in \Omega_1, \theta_2 \geqq \varrho\}$. The same objections hold as in Section 6. Other distance functions provide other results. If for example A_ϱ would have been defined by the inequality $\nu_1\theta_1 + \nu_2\theta_2 \geqq \varrho$ then the test which rejects for large values of $\nu_1 X_1 + \nu_2 X_2$ (which is a S.M.P. size-α test) would be "optimal". In more complicated testing problems with a restricted alternative (see [12]) the construction of such sub-alternatives seems to be a very arbitrary approach: either a particular S.M.P. test is shown to be "optimal" (and then M.S.S.M.P.(D) is more reasonable in our opinion), or it is extremely difficult to obtain the optimal test for a certain value of ϱ (and another value of ϱ provides another optimal test).

COMPARING THE TESTS IN THE VAN ZWET-OOSTERHOFF CASE $\psi = \pi/4$; $\alpha = .05$. In this section we tried to summarize the relevant results of [11], [12] and [13] and we introduced test φ_1. We shall have to compare the power-properties of φ_1 with those of φ^*, φ_0 and φ_2. We restrict the attention to the $\psi = \pi/4$ case because then φ_1 can be obtained relatively easily. In fig. 4 the critical regions of φ^*, φ_0 and φ_1 are drawn. In fig. 5 we plotted the maximum of the shortcoming over the half-line $l(\psi)$ (see fig. 4) as a function of ψ. We borrowed the graph for test φ_2 from OOSTERHOFF's results on [11] p. 107. Graphs for Fisher's test φ_3 have not been presented in the figs. 4 and 5 because these graphs are very close to those of φ_1. The results for φ^* and φ_0 were obtained from [13] though these results

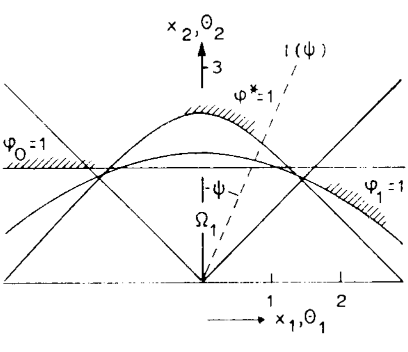

Fig. 4.

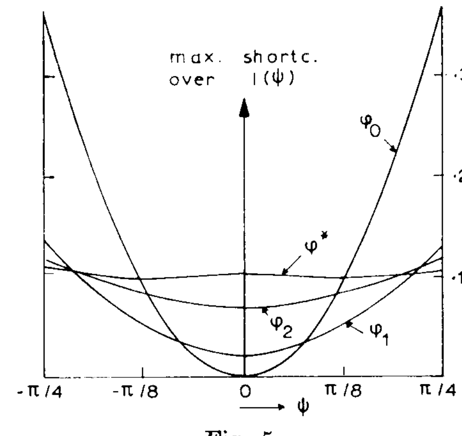

Fig. 5.

are also present in [11]. We conclude that from a common sense over-all point of view φ_1 and φ_3 are most attractive.

Problems from practice. See BARTHOLOMEW [1] who applied the likelihood-ratio principle, and [12] Part II where M.S.S.M.P.(D) tests were constructed. The latter results are attractive from the practical point of view but it is very hard to decide for which problems completely satisfactory power-properties are obtained. We have always been rather optimistic on this subject (see [12] and [13]) and sometimes too optimistic as was shown by Oosterhoff and in fig. 5.

8. $\boldsymbol{\Omega = \omega_6(\psi) \cup \omega_7(\psi)}$. First consider testing $H: \theta \in \Omega_0 = \omega_7(\psi)$ against $A: \theta \in \Omega_1 = \omega_6(\psi)$. If the test φ^* (respectively φ_1) of Section 7 is of size-α for testing this extended hypothesis $H: \theta \in \omega_7(\psi)$ (this will be the case if and only if ψ is sufficiently small; see Section 6, where φ_1 obviously is not of size-α for the extended hypothesis), then φ^* (respectively φ_1) is automatically M.S. size-α (respectively strongest w.r. λ) for our new problem. The test φ_0 is M.S.S.M.P. size-α. The likelihood-ratio principle does not provide the same test φ_2 as in Section 7.

In many situations the asymmetric above-mentioned Neyman-Pearson approach is less attractive than the approach where $H: \theta \in \Omega_0 = \omega_0$ has to be tested against the two-sided alternative $A: \theta \in \Omega_1 = \omega_6(\psi) \cup \omega_7(\psi)$. Of course the test which rejects if and only if $|x_2| \geqq u_{\frac{1}{2}\alpha}$ is M.S.S.M.P. unbiased size-α (see [12]). This test admits a simple interpretation: for $x_2 \geqq u_{\frac{1}{2}\alpha}$ one will state "$\theta \in \omega_6(\psi)$"; for $x_2 \leqq -u_{\frac{1}{2}\alpha}$ that "$\theta \in \omega_7(\psi)$". The other approaches lead to considerable difficulties: the critical regions will not fall apart into two different parts. A consistent 3-decision approach to these problems was given in [12] where minimax regret somewhere minimum risk unbiased procedures were derived.

(To be continued)